Procédé et système d'analyse de signaux vocaux pour la représentation compacte de locuteurs.

La présente invention concerne un procédé et un dispositif d'analyse de signaux vocaux.

L'analyse de signaux vocaux nécessite notamment de pouvoir représenter un locuteur. La représentation d'un locuteur par un mélange de gaussiennes ("Gaussian Mixture Model" ou GMM) est une représentation efficace de l'identité acoustique ou vocale d'un locuteur. Selon cette technique, il s'agit de représenter le locuteur, dans un espace acoustique de référence d'une dimension prédéterminée, par une somme pondérée d'un nombre prédéterminé de gaussiennes.

Ce type de représentation est précis lorsque l'on dispose d'un grand nombre de données, et qu'il n'y a pas de contraintes physiques pour stocker les paramètres du modèle, ni pour exécuter des calculs sur ces nombreux paramètres.

Or, en pratique, pour représenter un locuteur au sein de systèmes informatiques, il arrive que le temps de parole d'un locuteur soit court, et que la taille de la mémoire nécessaire à ces représentations, ainsi que les temps de calculs sur ces paramètres soient trop importants.

Il est donc important de chercher à représenter un locuteur de manière à réduire drastiquement le nombre de paramètres nécessaires à sa représentation tout en gardant des performances correctes. On entend par performance le taux d'erreurs de séquences vocales non reconnues comme appartenant ou non à un locuteur par rapport au nombre total de séquences vocales.

Des solutions en ce sens ont été proposées, notamment dans le document "SPEAKER INDEXING IN LARGE AUDIO DATABASES USING ANCHOR MODELS" par D.E. Sturim, D.A. Reynolds, E. Singer and J.P. Campbell. En effet, les auteurs proposent de représenter un locuteur, non plus de manière absolue

dans un espace acoustique de référence, mais de manière relative par rapport à un ensemble prédéterminé de représentations de locuteurs de référence appelés également modèles d'ancrages, pour lesquels on dispose de modèles GMM-UBM (UBM pour "Universal Background Model"). On évalue la proximité entre un locuteur et les locuteurs de référence au moyen d'une distance euclidienne. Cela diminue énormément les charges de calcul, mais les performances sont encore limitées et insuffisantes.

Au vu de ce qui précède, l'invention a pour but d'analyser des signaux vocaux en représentant les locuteurs par rapport à un ensemble prédéterminé de locuteurs de référence, avec un nombre de paramètre réduits diminuant les charges de calculs pour des application en temps réel, avec des performances acceptables, en comparaison d'une analyse utilisant une représentation par le modèle GMM-UBM.

On peut alors par exemple effectuer des indexations de documents audio de grandes bases de données où le locuteur est la clé d'indexation.

Ainsi, selon un aspect de l'invention, il est proposé un procédé d'analyse de signaux vocaux d'un locuteur (λ), utilisant une densité de probabilité représentant les ressemblances entre une représentation vocale du locuteur (λ) dans un modèle prédéterminé et un ensemble prédéterminé de représentations vocales d'un nombre E de locuteurs de référence dans ledit modèle prédéterminé, et on analyse la densité de probabilité pour en déduire des informations portant sur les signaux vocaux.

Cela permet de diminuer drastiquement le nombre de paramètres utilisés, et permet à des dispositifs mettant en œuvre ce procédé de pouvoir travailler en temps réel, en diminuant le temps de calcul, en diminuant la taille de la mémoire nécessaire.

Dans un mode de mise en œuvre préféré, on prend comme modèle prédéterminé un modèle absolu (GMM), de dimension D, utilisant un mélange de M gaussiennes pour lequel le locuteur (λ) est représenté par un ensemble de paramètres comprenant des

coefficients de pondération (α_i, i=1 à M) du mélange de gaussiennes dans ledit modèle absolu (GMM), des vecteurs de moyenne (μ_i, i=1 à M) de dimension D et des matrices de covariance (Σ_i, i=1 à M) de dimension D×D.

Dans un mode de mise en œuvre avantageux, on représente la densité de probabilité des ressemblances entre la représentation desdits signaux vocaux du locuteur (λ) et l'ensemble prédéterminé de représentations vocales des locuteurs de référence par une distribution gaussienne ($\psi(\mu^\lambda, \Sigma^\lambda)$) de vecteur de moyenne (μ^λ) de dimension E et de matrice de covariance (Σ^λ) de dimension E×E estimés dans l'espace des ressemblances à l'ensemble prédéterminé des E locuteurs de référence.

Dans un mode de mise en œuvre préféré, l'on définit la ressemblance ($\psi(\mu^\lambda, \Sigma^\lambda)$) du locuteur ($\lambda$) par rapport aux E locuteurs de référence, locuteur (λ) pour lequel on dispose de N_λ segments de signaux vocaux représentés par N_λ vecteurs de l'espace des ressemblances par rapport à l'ensemble prédéterminé des E locuteurs de référence, en fonction d'un vecteur de moyenne (μ^λ) de dimension E et d'une matrice de covariance (Σ^λ) des ressemblances du locuteur (λ) par rapport aux E locuteurs de référence.

Dans un mode de mise en œuvre avantageux, on introduit en outre des informations à priori dans les densités de probabilité des ressemblances ($\psi(\tilde{\mu}^\lambda, \tilde{\Sigma}^\lambda)$) par rapport aux E locuteurs de référence.

Dans un mode de mise en œuvre préféré, la matrice de covariance du locuteur (λ) est indépendante dudit locuteur ($\tilde{\Sigma}^\lambda = \tilde{\Sigma}$).

Selon un autre aspect de l'invention, il est proposé un système d'analyse de signaux vocaux d'un locuteur (λ), comprenant des bases de données dans lesquelles sont stockés des signaux vocaux d'un ensemble prédéterminé de E locuteurs de référence et leurs représentations vocales associées dans un

modèle prédéterminé, ainsi que des bases de données d'archives audio, caractérisé en ce qu'il comprend des moyens d'analyse des signaux vocaux utilisant une représentation vectorielle des ressemblances entre la représentation vocale du locuteur et l'ensemble prédéterminé de représentations vocales de E locuteurs de référence.

Dans un mode de réalisation avantageux, les bases de données mémorisent également l'analyse des signaux vocaux effectuée par lesdits moyens d'analyse.

L'invention peut s'appliquer à l'indexation de documents audio, toutefois d'autres applications peuvent également être envisagées, telles que l'identification acoustique d'un locuteur ou la vérification de l'identité d'un locuteur.

D'autres buts, caractéristiques et avantages de l'invention apparaîtront à la lecture de la description suivante, donnée à titre d'exemple non limitatif, et faite en référence à l'unique dessin annexé illustrant une mise en application d'une utilisation du procédé pour l'indexation de documents audio.

La figure représente une application du système selon un aspect de l'invention pour l'indexation de bases de données audio. Bien entendu, l'invention s'applique également à l'identification acoustique d'un locuteur ou la vérification de l'identité d'un locuteur, c'est-à-dire, de manière générale, à la reconnaissance d'informations relatives au locuteur dans le signal acoustique. Le système comprend un moyen pour recevoir des données vocales d'un locuteur, par exemple un micro 1, relié par une connexion 2 avec ou sans fil à des moyens d'enregistrement 3 d'une requête énoncée par un locuteur λ et comprenant un ensemble de signaux vocaux. Les moyens d'enregistrement 3 sont reliés par une connexion 4 à des moyens de stockage 5 et, par une connexion 6, à des moyens de traitement acoustique 7 de la requête. Ces moyens de traitement acoustiques transforment les signaux vocaux du locuteur λ en une représentation dans un espace acoustique de dimension D par un modèle GMM de représentation du locuteur λ.

Cette représentation est définie par une somme pondérée de M gaussiennes selon les équations :

$$\begin{cases} p(x|\lambda) = \sum_{i=1}^{M} \alpha_i b_i(x) & (1) \\ b_i(x) = \frac{1}{(2\pi)^{D/2} \cdot |\Sigma_i|^{1/2}} \times \exp\left[-\frac{1}{2} \, {}^t(x-\mu_i) \, \Sigma_i^{-1} \, (x-\mu_i) \right] & (2) \\ \sum_{i=1}^{M} \alpha_i = 1 & (3) \end{cases}$$

dans lesquelles :

D est la dimension de l'espace acoustique du modèle GMM absolu;

x est un vecteur acoustique de dimension D, ie vecteur des coefficients cepstraux d'une séquence de signal vocal du locuteur λ dans le modèle GMM absolu ;

M désigne le nombre de gaussiennes du modèle GMM absolu, généralement puissance de 2 comprise entre 16 et 1024 ;

$b_i(x)$ désigne, pour i=1 à D, densités gaussiennes, paramétrées par un vecteur de moyenne μ_i de dimension D et une matrice de covariance Σ_i de dimension D×D; et

α_i désigne, pour i=1 à D représentent les coefficients de pondération du mélange de gaussiennes dans le modèle GMM absolu.

Les moyens de traitement acoustique 7 de la requête sont reliés par une connexion 8 à des moyens d'analyse 9. Ces moyens d'analyse 9 sont aptes à représenter un locuteur par un vecteur de densité de probabilité représentant les ressemblances entre la représentation vocale dudit locuteur dans le modèle GMM choisi et des représentations vocales de E locuteurs de référence dans le modèle GMM choisi. Les moyens d'analyse 9 sont en outre aptes à effectuer des tests de vérification et/ou d'identification d'un locuteur.

Pour réaliser ces tests, les moyens d'analyse procèdent à l'élaboration du vecteur de densités de probabilités, c'est-à-dire des ressemblances entre le locuteur et les locuteurs de référence.

Il s'agit de décrire une représentation pertinente d'un seul segment x du signal du locuteur λ au moyen des équations suivantes :

$$\begin{cases} w^{\lambda} = \begin{pmatrix} \tilde{p}\left(x^{\lambda}|\overline{\lambda}_1\right) \\ \vdots \\ \tilde{p}\left(x^{\lambda}|\overline{\lambda}_E\right) \end{cases} & (4) \\ \tilde{p}\left(x^{\lambda}|\overline{\lambda}_j\right) = \frac{1}{T_x}\log\left(\frac{p\left(x^{\lambda}|\overline{\lambda}_j\right)}{p\left(x^{\lambda}|\overline{\lambda}_{UBM}\right)}\right) & (5) \\ p\left(x|\overline{\lambda}\right) = \sum_{k=1}^{M}\alpha_k b_k(x) \quad \text{où} \quad \sum_{k=1}^{M}\alpha_k = 1 & (6) \\ b_k(x) = \frac{1}{(2\pi)^{D/2}|\Sigma_k|^{1/2}} \times \exp\left[-\frac{1}{2}\,{}^t(x-\mu_k)(\Sigma_k)^{-1}(x-\mu_k)\right] & (7) \end{cases}$$

dans lesquelles :

w^{λ} est un vecteur de l'espace des ressemblances à l'ensemble prédéterminé des E locuteurs de référence représentant le segment x dans cet espace de représentation ;

$\tilde{p}\left(x^{\lambda}|\overline{\lambda}_j\right)$ est une densité de probabilité ou probabilité normalisée par un modèle universel, représentant la ressemblance de la représentation acoustique x^{λ} d'un segment de signal vocal d'un locuteur λ, sachant un locuteur de référence $\overline{\lambda}_j$;

T_x est le nombre de trames ou de vecteurs acoustiques du segment de parole x ;

$p\left(x^{\lambda}|\overline{\lambda}_j\right)$ est une probabilité représentant la ressemblance de la représentation acoustique x^{λ} d'un segment de signal vocal d'un locuteur λ, sachant un locuteur de référence $\overline{\lambda}_j$;

$p\left(x^{\lambda}|\overline{\lambda}_{UBM}\right)$ est une probabilité représentant la ressemblance de la représentation acoustique x^{λ} d'un segment de signal vocal d'un locuteur λ dans le modèle du monde UBM;

M est le nombre de gaussiennes du modèle GMM relatif, généralement puissance de 2 comprise entre 16 et 1024 ;

D est la dimension de l'espace acoustique du modèle GMM absolu;

x^{λ} est un vecteur acoustique de dimension D, ie vecteur des coefficients cepstraux d'une séquence de signal vocal du locuteur λ dans le modèle GMM absolu;

$b_k(x)$ représente, pour k=1 à D, des densités gaussiennes, paramétrées par un vecteur de moyenne μ_k de dimension D et une matrice de covariance Σ_k de dimension D×D ;

α_k représente, pour k=1 à D, les coefficients de pondération du mélange de gaussiennes dans le modèle GMM absolu ;

A partir des représentations W_j des segments de parole x_j ($j=1,..,N_{\lambda}$) du locuteur λ, on représente le locuteur λ par la distribution gaussienne ψ de paramètres μ^{λ} et Σ_{λ} définis par les relations suivantes:

$$\begin{cases} \mu^{\lambda} = \left\{\mu_i^{\lambda}\right\}_{i=1,..,E} \quad \text{avec} \quad \mu_i^{\lambda} = \dfrac{1}{N_{\lambda}} \displaystyle\sum_{j=1}^{N_{\lambda}} \tilde{p}\left(x_j^{\lambda}|\overline{\lambda}_i\right) & (8) \\ \Sigma^{\lambda} = \left\{\Sigma_{ii'}^{\lambda}\right\}_{i,i'=1,..,E} \quad \text{avec} \quad \Sigma_{ii'}^{\lambda} = \dfrac{1}{N_{\lambda}} \displaystyle\sum_{j=1}^{N_{\lambda}} \left(\tilde{p}\left(x_j^{\lambda}|\overline{\lambda}_i\right) - \mu_i^{\lambda}\right)\left(\tilde{p}\left(x_j^{\lambda}|\lambda_{i'}\right) - \mu_{i'}^{\lambda}\right) & (9) \end{cases}$$

dans lesquelles μ_i^{λ} représente des composantes du vecteur de moyenne μ^{λ} de dimension E des ressemblances $\psi\left(\mu^{\lambda},\Sigma^{\lambda}\right)$ du locuteur λ par rapport aux E locuteurs de référence, et $\Sigma_{ii'}^{\lambda}$ représente des composantes de la matrice de covariance Σ^{λ} de dimension E×E des ressembles $\psi\left(\mu^{\lambda},\Sigma^{\lambda}\right)$ du locuteur λ par rapport aux E locuteurs de référence.

Les moyens d'analyse 9 sont reliés par une connexion 10 à des moyens d'apprentissage 11 permettant de calculer les représentations vocales, sous forme de vecteurs de dimension D,

des E locuteurs de référence dans le modèle GMM choisi. Les moyens d'apprentissage 11 sont reliés par une connexion 12 à une base de données 13 comprenant des signaux vocaux d'un ensemble prédéterminé de locuteurs et leurs représentations vocales associées dans le modèle GMM de référence. Cette base de données peut également stocker le résultat de l'analyse de signaux vocaux de locuteurs initiaux excepté lesdits E locuteurs de référence. La base de données 13 est reliée par la connexion 14 aux moyens d'analyse 9 et par une connexion 15 aux moyens de traitement acoustique 7.

Le système comprend en outre une base de données 16 reliée par une connexion 17 aux moyens de traitement acoustique 7, et par une connexion 18 aux moyens d'analyse 9. La base de données 16 comprend des archives audio sous formes d'articles vocaux, ainsi que les représentations vocales associées dans le modèle GMM choisi. La base de données 16 est également apte à stocker les représentations associées des articles audio calculées par les moyens d'analyse 9. Les moyens d'apprentissage 11 sont en outre reliés par une connexion 19 aux moyens de traitement acoustique 7.

On va maintenant décrire un exemple de fonctionnement de ce système pouvant fonctionner en temps réel car le nombre de paramètres utilisés est nettement réduit par rapport au modèle GMM, et car beaucoup d'étapes peuvent être effectuées hors-ligne.

Le module d'apprentissage 11 va déterminer les représentations dans le modèle GMM de référence des E locuteurs de référence au moyen des signaux vocaux de ces E locuteurs de référence stockés dans la base de données 13, et des moyens de traitement acoustique 7. Cette détermination s'effectue selon les relations (1) à (3) mentionnées ci-dessus. Cet ensemble de E locuteurs de référence va représenter le nouvel espace de représentation acoustique. Ces représentations des E locuteurs de référence dans le modèle GMM sont stockées en mémoire, par

exemple dans la base de données 13. Tout cela peut être effectué hors-ligne.

Lorsque l'on reçoit des données vocales d'un locuteur λ, par exemple par le micro 1, celles-ci sont transmises par la connexion 2 aux moyens d'enregistrement 3 aptes à effectuer le stockage de ces données dans les moyens de stockage 5 à l'aide de la connexion 4. Les moyens d'enregistrement 3 transmettent cet enregistrement aux moyens de traitement acoustique 7 par la connexion 6. Les moyens de traitement acoustique 7 calculent une représentation vocale du locuteur dans le modèle GMM prédéterminé comme exposé précédemment en référence aux relations (1) à (3) ci-dessus.

En outre, les moyens de traitement acoustique 7 ont calculé, par exemple hors-ligne, les représentations vocales d'un ensemble de S locuteurs de test et d'un ensemble de T locuteurs dans le modèle GMM prédéterminé. Ces ensembles sont distincts. Ces représentations sont stockées dans la base de données 13. Les moyens d'analyse 9 calculent, par exemple hors-ligne, une représentation vocale des S locuteurs et des T locuteurs par rapport aux E locuteurs de référence. Cette représentation est une représentation vectorielle par rapport à ces E locuteurs de référence, comme décrit précédemment. Les moyens d'analyse 9 effectuent également, par exemple hors-ligne, une représentation vocale des S locuteurs et des T locuteurs par rapport aux E locuteurs de référence, et une représentation vocale des articles des locuteurs de la base audio. Cette représentation est une représentation vectorielle par rapport à ces E locuteurs de référence.

Les moyens de traitement 7 transmettent la représentation vocale du locuteur λ dans le modèle GMM prédéterminé aux moyens d'analyse 9, qui calculent une représentation vocale du locuteur λ. Cette représentation est une représentation par densité de probabilité des ressemblances aux E locuteurs de référence. Elle est calculée en introduisant de l'information à priori au

moyen des représentations vocales de T locuteurs. En effet, l'utilisation de cette information à priori permet de garder une estimation fiable, même lorsque le nombre de segments de paroles disponibles du locuteur λ est faible. On introduit de l'information à priori au moyen des équations suivantes :

$$\tilde{\mu}^{\lambda} = \frac{N_0 \mu_0 + N_\lambda \mu^{\lambda}}{N_0 + N_\lambda} \qquad (10)$$

$$W = \left(w_1^{loc\_1} \ldots w_{N_1}^{loc\_1} \ldots w_1^{loc\_T} \ldots w_{N_T}^{loc\_T} \right) \qquad (11)$$

dans lesquelles :

μ^{λ} : vecteur de moyenne de dimension E des ressemblances $\psi(\mu^{\lambda}, \Sigma^{\lambda})$ du locuteur λ par rapport aux E locuteurs de référence ;

N_λ : nombre de segments de signaux vocaux du locuteur λ représentés par N_λ vecteurs de l'espace des ressemblances à l'ensemble prédéterminé des E locuteurs de référence ;

W : matrice de toutes les données initiales d'un ensemble de T locuteurs loc_i, pour i=1 à T, dont les colonnes sont des vecteurs de dimension E représentant un segment de signal vocal représenté par un vecteur de l'espace des ressemblances à l'ensemble prédéterminé des E locuteurs de référence, chaque locuteur loc_i ayant N_i segments vocaux, caractérisé par son vecteur de moyennes μ_0 de dimension E, et par sa matrice de covariance Σ_0 de dimension E×E ;

$\tilde{\mu}^{\lambda}$: vecteur de moyenne de dimension E des ressemblances $\psi(\tilde{\mu}^{\lambda}, \tilde{\Sigma}^{\lambda})$ du locuteur λ par rapport aux E locuteurs de référence, avec introduction d'informations à priori; et

$\tilde{\Sigma}^{\lambda}$: matrice de covariance de dimension E×E des ressemblances $\psi(\tilde{\mu}^{\lambda}, \tilde{\Sigma}^{\lambda})$ du locuteur λ par rapport aux E locuteurs de référence avec introduction d'informations à priori.

On peut prendre de surcroît une unique matrice de covariance pour chaque locuteur, ce qui permet d'orthogonaliser ladite matrice hors-ligne, et les calculs de densités de probabilités

seront alors effectués avec des matrices de covariance diagonales. Dans ce cas, cette unique matrice de covariance est définie selon les relations :

$$\begin{cases} \tilde{\Sigma}_{ii'} = \frac{1}{N_0} \sum_{s=1}^{T} \sum_{j \in I_s} \left(W_{ij} - \overline{W}_{is}\right)\left(W_{i'j} - \overline{W}_{i's}\right) & (12) \\ \overline{W}_{is} = \frac{1}{N_T} \sum_{j \in I_s} W_{ij} & (13) \end{cases}$$

dans lesquelles

W est une matrice de toutes les données initiales d'un ensemble de T locuteurs loc_i, pour i=1 à T, dont les colonnes sont des vecteurs de dimension E représentant un segment de signal vocal représenté par un vecteur de l'espace des ressemblances à l'ensemble prédéterminé des E locuteurs de référence, chaque locuteur loc_i ayant N_i segments vocaux, caractérisé par son vecteur de moyennes μ_0 de dimension E, et par sa matrice de covariance Σ_0 de dimension E×E.

Ensuite les moyens d'analyse 9 vont comparer les représentations vocales de la requête et des articles de la base articles de la base par des tests en identification et/ou vérification du locuteurs. Le test en identification de locuteur consiste à évaluer une mesure de vraisemblance entre le vecteur du segment de test w_x et l'ensemble des représentations des articles de la base audio. Le locuteur identifié correspond à celui qui donne un score de vraisemblance maximal, soit $\hat{\lambda} = \arg\max_{\lambda} p\left(w_x \middle| \tilde{\mu}^{\lambda}, \tilde{\Sigma}^{\lambda}\right)$ (14) parmi l'ensemble des S locuteurs.

Le test en vérification de locuteur consiste à calculer un score de vraisemblance entre le vecteur du segment de test w_x et l'ensemble des représentations des articles de la base audio normalisé par son score de vraisemblance avec la représentation de l'information à priori. Le segment est authentifié si le score excède un seuil donné prédéterminé, ledit score étant donné par la relation suivante:

$$\text{score} = \frac{p\left(w_x \middle| \tilde{\mu}^{\lambda}, \tilde{\Sigma}^{\lambda}\right)}{p\left(w_x \middle| \mu_0, \Sigma_0\right)} \qquad (15)$$

Chaque fois que le locuteur λ est reconnu dans un article de la base, on indexe cet article au moyen d'une information permettant de savoir que le locuteur λ parle dans cet article audio.

On peut également appliquer cette invention à d'autres utilisations, comme la reconnaissance ou l'identification d'un locuteur.

Cette représentation compacte d'un locuteur permet de réduire de façon drastique le coût de calcul, car il y a beaucoup moins d'opération élémentaires au vu de la réduction drastique du nombre de paramètres nécessaires à la représentation d'un locuteur.

Par exemple, pour une requête de 4 secondes de paroles d'un locuteur, c'est-à-dire 250 trames, pour un modèle GMM de dimension 27, à 16 gaussiennes le nombre d'opérations élémentaires est réduit d'un facteur 540, ce qui réduit énormément le temps calcul. En outre, la taille de mémoire utilisée pour stocker les représentations des locuteurs est nettement réduite.

L'invention permet donc d'analyser des signaux vocaux d'un locuteur en réduisant de manière drastique le temps de calcul et la taille mémoire de stockage des représentations vocales des locuteurs.

REVENDICATIONS

1. Procédé d'analyse de signaux vocaux d'un locuteur (λ), caractérisé en ce que l'on utilise une densité de probabilité représentant les ressemblances entre une représentation vocale du locuteur (λ) dans un modèle prédéterminé et un ensemble prédéterminé de représentations vocales d'un nombre E de locuteurs de référence dans ledit modèle prédéterminé, et on analyse la densité de probabilité pour en déduire des informations sur les signaux vocaux.

2. Procédé selon la revendication 1, caractérisé en ce que l'on prend comme modèle prédéterminé un modèle absolu (GMM), de dimension D, utilisant un mélange de M gaussiennes pour lequel le locuteur (λ) est représenté par un ensemble de paramètres comprenant des coefficients de pondération (α_i, i=1 à M) du mélange de gaussiennes dans ledit modèle absolu (GMM), des vecteurs de moyenne (μ_i, i=1 à M) de dimension D et des matrices de covariance (Σ_i, i=1 à M) de dimension D×D.

3. Procédé selon la revendication 2, caractérisé en ce que l'on représente la densité de probabilité des ressemblances entre la représentation desdits signaux vocaux du locuteur (λ) et l'ensemble prédéterminé de représentations vocales des locuteurs de référence par une distribution gaussienne ($\psi(\mu^\lambda, \Sigma^\lambda)$) de vecteur de moyenne (μ^λ) de dimension E et de matrice de covariance (Σ^λ) de dimension E×E estimés dans l'espace des ressemblances à l'ensemble prédéterminé des E locuteurs de référence.

4. Procédé selon la revendication 3, caractérisé en ce que l'on définit la ressemblance ($\psi(\mu^\lambda, \Sigma^\lambda)$) du locuteur ($\lambda$) par rapport aux E locuteurs de référence, locuteur (λ) pour lequel on dispose de N_λ segments de signaux vocaux représentés par N_λ vecteurs de l'espace des ressemblances par rapport à l'ensemble prédéterminé des E locuteurs de référence, en fonction d'un

vecteur de moyenne (μ^{λ}) de dimension E et d'une matrice de covariance (Σ^{λ}) des ressemblances du locuteur (λ) par rapport aux E locuteurs de référence.

5. Procédé selon la revendication 4, caractérisé en ce que l'on introduit en outre des informations à priori dans les densités de probabilité des ressemblances ($\psi(\tilde{\mu}^{\lambda},\tilde{\Sigma}^{\lambda})$) par rapport aux E locuteurs de référence.

6. Procédé selon la revendication 5, caractérisé en ce que la matrice de covariance du locuteur (λ) est indépendante dudit locuteur ($\tilde{\Sigma}^{\lambda}=\tilde{\Sigma}$).

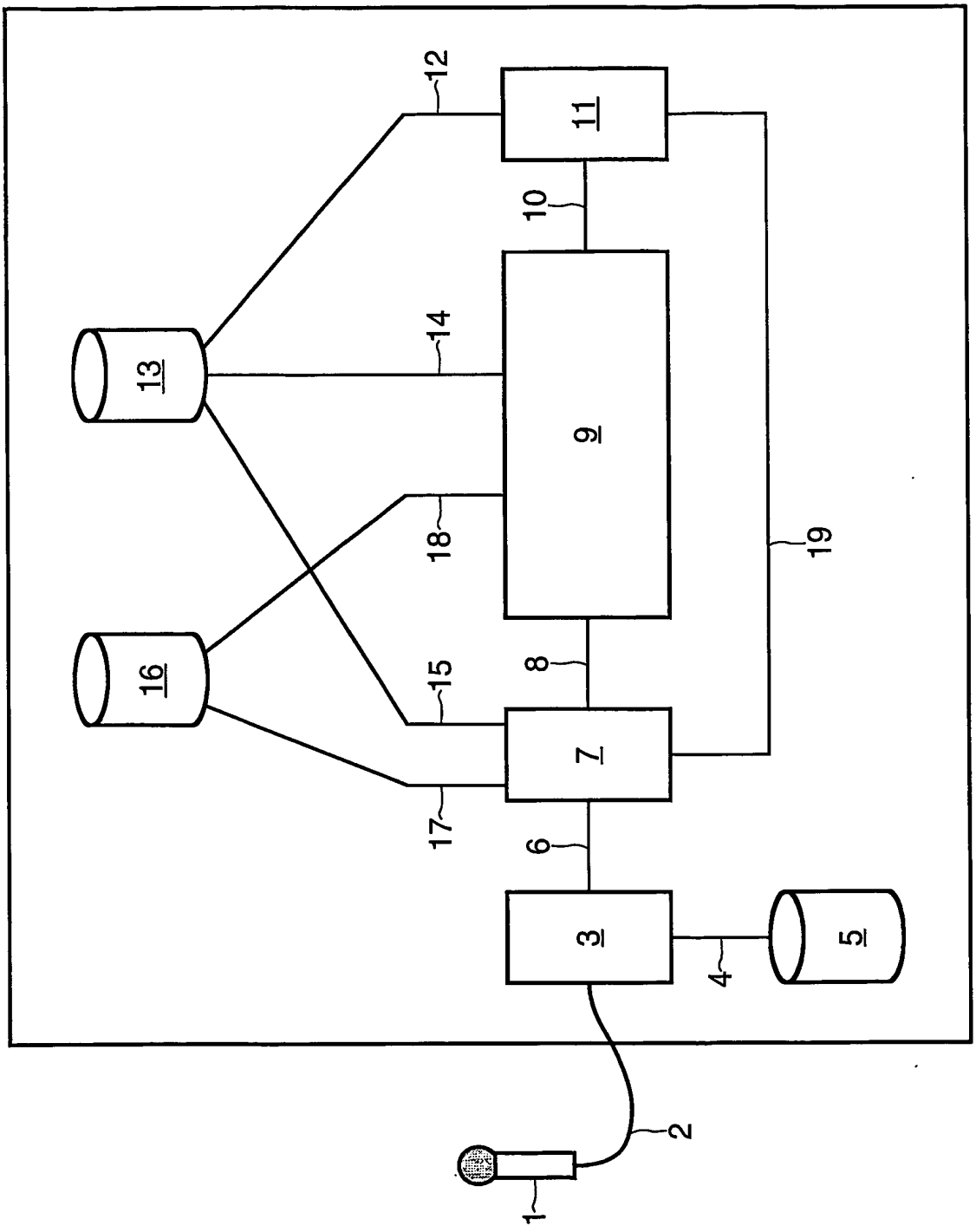
7. Système d'analyse de signaux vocaux d'un locuteur (λ), comprenant des bases de données dans lesquelles sont stockés des signaux vocaux d'un ensemble prédéterminé de locuteurs et leurs représentations vocales associées dans un modèle prédéterminé par mélange de gaussiennes, ainsi que des bases de données d'archives audio, caractérisé en ce qu'il comprend des moyens d'analyse des signaux vocaux utilisant une représentation vectorielle des ressemblances entre la représentation vocale du locuteur (λ) et l'ensemble prédéterminé de représentations vocales de E locuteurs de référence.

8. Système selon la revendication 7, caractérisé en ce que les bases de données mémorisent également l'analyse des signaux vocaux effectuée par lesdits moyens d'analyse.

9. Utilisation d'un procédé selon l'une quelconque des revendications 1 à 6, pour une indexation de documents audio.

10. Utilisation d'un procédé selon l'une quelconque des revendications 1 à 6, pour une identification d'un locuteur.

11. Utilisation d'un procédé selon l'une quelconque des revendications 1 à 6, pour une vérification d'un locuteur.


12
11
10
14
13
9
18
19
16
15
8
7
17
6
3
5
4
2
1

INTERNATIONAL SEARCH REPORT

International Application No
PCT/FR 03/02037

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L17/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, INSPEC, WPI Data, PAJ, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	STURIM D E ET AL: "Speaker indexing in large audio databases using anchor models" 2001 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS (CAT. NO.01CH37221), 2001 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS, SALT LAKE CITY, UT, USA, 7-11 MAY 2001, pages 429-432 vol.1, XP002272038 2001, Piscataway, NJ, USA, IEEE, USA ISBN: 0-7803-7041-4 abstract page 429-430, paragraph 2.0; figure 1 --- -/--	1-11

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
3 March 2004	25/03/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Greiser, N

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/FR 03/02037

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	REYNOLDS D A: "Speaker identification and verification using Gaussian mixture speaker models" SPEECH COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 17, no. 1, 1 August 1995 (1995-08-01), pages 91-108, XP004062392 ISSN: 0167-6393 abstract page 94, paragraph 2.0 ---	1,7
A	US 6 411 930 B1 (BURGES CHRISTOPHER JOHN) 25 June 2002 (2002-06-25) column 1, line 10-27 column 2, line 52 -column 3, line 46 -----	1,7

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/FR 03/02037

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 6411930 B1	25-06-2002	NONE	

RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale No
PCT/FR 03/02037

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
CIB 7 G10L17/00

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)
CIB 7 G10L

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de donnees électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)
EPO-Internal, INSPEC, WPI Data, PAJ, IBM-TDB

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie °	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	STURIM D E ET AL: "Speaker indexing in large audio databases using anchor models" 2001 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS (CAT. NO.01CH37221), 2001 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS, SALT LAKE CITY, UT, USA, 7-11 MAY 2001, pages 429-432 vol.1, XP002272038 2001, Piscataway, NJ, USA, IEEE, USA ISBN: 0-7803-7041-4 abrégé page 429-430, alinéa 2.0; figure 1 --- -/--	1-11

[X] Voir la suite du cadre C pour la fin de la liste des documents

[X] Les documents de familles de brevets sont indiqués en annexe

° Catégories spéciales de documents cités:

"A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent

"E" document antérieur, mais publié à la date de dépôt international ou après cette date

"L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)

"O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens

"P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

"X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

"Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

"&" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée	Date d'expédition du présent rapport de recherche internationale
3 mars 2004	25/03/2004

Nom et adresse postale de l'administration chargée de la recherche internationale	Fonctionnaire autorisé
Office Européen des Brevets, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Greiser, N

RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale No
PCT/FR 03/02037

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	REYNOLDS D A: "Speaker identification and verification using Gaussian mixture speaker models" SPEECH COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 17, no. 1, 1 août 1995 (1995-08-01), pages 91-108, XP004062392 ISSN: 0167-6393 abrégé page 94, alinéa 2.0 ---	1,7
A	US 6 411 930 B1 (BURGES CHRISTOPHER JOHN) 25 juin 2002 (2002-06-25) colonne 1, ligne 10-27 colonne 2, ligne 52 -colonne 3, ligne 46 -----	1,7

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demande internationale No
PCT/FR 03/02037

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
US 6411930 B1	25-06-2002	AUCUN	